

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-149485

(43)Date of publication of application : 27.05.1994

(51)Int.Cl.

G06F 3/06
G06F 12/00

(21)Application number : 04-296766

(71)Applicant : FUJITSU LTD

(22)Date of filing : 06.11.1992

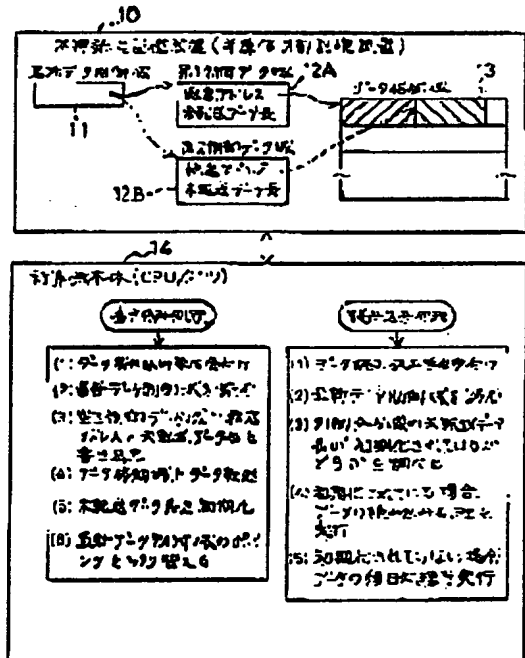
(72)Inventor : SHIGA KOICHI

(54) DATA COMPLETION GUARANTEE PROCESSING METHOD

(57)Abstract:

PURPOSE: To provide an efficient completion guarantee method for which a CPU overhead is small concerning a data completion guarantee processing method for guaranteeing the completion of data during transfer to a nonvolatile storage device such as a semiconductor external storage device or the like.

CONSTITUTION: Plural control data areas 12A and 12B for guaranteeing the completion of a data group are prepared on the nonvolatile storage device 10 and a latest data control area 11 for pointing the control data area for holding latest data for which the write of the data group is completed within the control data areas is provided on the nonvolatile storage device 10. Then, in the case of updating the control data areas, the data group to be updated is written in the control data area not pointed from the latest data control area 11 and when the write of the data group is completed, a pointer in the latest data control area 11 is written so as to point the control data area for which the write is completed by an instruction by which the completion of data transfer is guaranteed.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平6-149485

(43) 公開日 平成6年(1994)5月27日

(51) Int.Cl.⁵G 0 6 F 3/06
12/00

識別記号

3 0 4 C 7165-5B
5 1 8 M 8526-5B

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数2(全13頁)

(21) 出願番号 特願平4-296766

(22) 出願日 平成4年(1992)11月6日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中1015番地

(72) 発明者 志賀 浩一

神奈川県川崎市中原区上小田中1015番地

富士通株式会社内

(74) 代理人 弁理士 小笠原 吉義 (外2名)

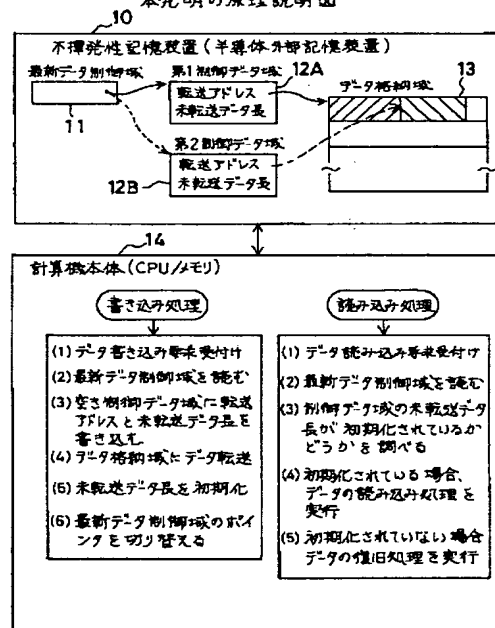
(54) 【発明の名称】 データ完結性保証処理方法

(57) 【要約】

【目的】半導体外部記憶装置などの不揮発性記憶装置10へ転送中のデータの完結性を保証するデータ完結性保証処理方法に関し、CPUオーバーヘッドの小さい効率よい完結性の保証方法を提供することを目的とする。

【構成】データ群の完結性を保証するための制御データ域12A、12Bを、不揮発性記憶装置10上に複数個用意し、その制御データ域のうち、データ群の書き込みが完了した最新のデータを保持する制御データ域をポイントする最新データ制御域11を、不揮発性記憶装置10上に設ける。そして、制御データ域を更新する場合に、最新データ制御域11からポイントされないほうの制御データ域に更新するデータ群を書き込み、そのデータ群の書き込みが完了した時点で、最新データ制御域11におけるポイントを、書き込みが完了した制御データ域をポイントするように、データ転送の完結性が保証される1命令で書き込む。

本発明の原理説明図



1

【特許請求の範囲】

【請求項1】 計算機本体(14)と、該計算機本体(14)の発行する命令によって読み書き可能な不揮発性の記憶装置(10)とを備え、前記記憶装置(10)上に設けられた制御データ域(12)に書き込むデータ群の完結性を保証する計算機システムにおけるデータ完結性保証処理方法において、前記制御データ域(12)を、前記記憶装置(10)上に複数個用意し、前記複数個の制御データ域(12)のうち、データ群の書き込みが完了した最新のデータを保持する制御データ域(12)をポイントする最新データ制御域(11)を、前記記憶装置(10)上に設け、制御データ域(12)を更新する場合に、前記最新データ制御域(11)からポイントされないほうの制御データ域(12)に更新するデータ群を書き込み、そのデータ群の書き込みが完了した時点で、前記最新データ制御域(11)におけるポイントを、書き込みが完了した制御データ域(12)をポイントするように、ポイント書き込み処理の完結性が保証される1命令で書き込み、前記最新データ制御域(11)からポイントされる制御データ域(12)を、有効なデータを保持する制御データ域(12)として扱うことを特徴とするデータ完結性保証処理方法。

【請求項2】 計算機本体(14)と、該計算機本体(14)の発行する命令によって読み書き可能な不揮発性の記憶装置(10)とを備え、前記記憶装置(10)上に設けられたデータ格納域(13)に対して転送するレコードの完結性を保証する計算機システムにおけるデータ完結性保証処理方法において、前記データ格納域(13)に転送されたレコードの完結性を検査するためのデータ群を格納する制御データ域(12)を、前記記憶装置(10)上に複数個用意し、前記複数個の制御データ域(12)のうち、データ群の書き込みが完了した最新のデータを保持する制御データ域(12)をポイントする最新データ制御域(11)を、前記記憶装置(10)上に設け、前記データ格納域(13)にレコードを転送する場合に、前記最新データ制御域(11)からポイントされないほうの制御データ域(12)にそのレコードの完結性を検査するためのデータ群を書き込むと共に、前記データ格納域(13)にレコードを転送し、それらが完了した時点で、前記最新データ制御域(11)におけるポイントを、書き込みが完了した制御データ域(12)をポイントするように、ポイント書き込み処理の完結性が保証される1命令で書き込み、前記データ格納域(13)に転送したレコードの完結性およびその完結性を検査するための制御データ域(12)のデータ群の完結性を保証することを特徴とするデータ完結性保証処理方法。

【発明の詳細な説明】

2

【0001】

【産業上の利用分野】本発明は、計算機システムにおいて、半導体外部記憶装置などの不揮発性記憶装置へ転送中のデータが、障害のため不定データとなることを回避し、データの完結性を保証するデータ完結性保証処理方法に関する。

【0002】複数の計算機を通信路で結合した形態のシステムを、複合システムと呼ぶ。複合システムを構成する計算機を、ここではクラスタと呼ぶ。このような複合システムにおいて、大量のトランザクションを処理するためには、不揮発性の半導体記憶装置を外部記憶媒体として使用することが有効である。

【0003】半導体を利用した高速入出力が可能な外部記憶媒体を、以後「システム記憶（SSU）」と称する。特に、システム記憶を利用してデータベースのシステム間共用制御を行う場合、ログバッファをシステム記憶に獲得して、このバッファをシステム間で共用するロギング方式がシステムの効率の上で有効である。

【0004】一方、システム記憶に転送されるデータは、本体系のハードウェア障害（例えば、CPU障害）によって、その完結性が保証されない場合がある。そのような転送データの完結性を効率よく保証する技術が必要とされる。

【0005】

【従来の技術】図8は、従来技術によるデータ転送方式を説明するための図である。記憶装置に転送中のデータが、CPU障害のようなハードウェア障害によって、転送前の値でもなく、転送後の値でもない状態になることがある。このような完結性が保証されないデータを不定データという。

【0006】従来、このような不定データを検出するために、図8に示すように、転送するデータ部の前後に、チェックバイトを配置することが行われていた。データ部の前後のチェックバイトに、例えばシーケンス番号やタイムスタンプなどをもとにした同じ値を設定してデータ転送を行う。次に、そのデータを読み込んだときに、前後のチェックバイトの値が異なっていれば、障害のため不定データになったと判断する。

【0007】図8に示すようなチェックバイトによるデータ転送方式では、データ転送が完結したか否か、すなわちデータが不定データであるか否かを判定することはできるが、転送データ自体の完結性を保証することはできなかった。

【0008】例えば、図9に示すように、クラスタ#1が、不揮発性記憶装置10にレコードaを書き込んでいたとする。このレコードaのデータを転送している途中で、クラスタ#1が、障害によりクラッシュしたとすると、不揮発性記憶装置10に書き込まれたデータは、レコードa'のように、レコードaの内容とも元の内容とも異なる値を持つデータになることがある。

3

【0009】このレコードa'をクラスタ#2が読み込んだとすると、図8に示すようなチェックバイトの検査により、レコードa'が不定データであることは検出できるが、正しいレコードaの内容または元の完結した内容を得ることはできない。

【0010】そのため、従来、不揮発性記憶装置10には、各クラスタ等が頻繁に使用する制御データなどを配置することができないという問題があった。すなわち、システム記憶上に配置した制御データ（例えば、クラスタ間排他情報など）が不定となった場合、その不定データを参照した後の複合システムの動作は保証されないため、システム記憶上に制御データを配置することができなかった。

【0011】この問題を解決するため、特開平3-62244号公報に示されるような半導体外部記憶装置におけるデータ保証制御方式の技術が提案されている。この方式では、半導体外部記憶装置への転送データを、外部不揮発記憶媒体上に退避するとともに、その転送データに関するリカバリ情報を外部不揮発記憶媒体上に設定する書き込み処理手段と、半導体外部記憶装置に対する書き込みまたは読み込みのアクセス要求の処理において、外部不揮発記憶媒体に設定されたりリカバリ情報を参照し、先のアクセスによる転送データが障害により不定データになっていることを検出した場合に、外部不揮発記憶媒体に退避した転送データに基づいて、不定データの復旧を行うリカバリ処理手段とを備えることにより、不定データを復旧することができるようにしている。

【0012】

【発明が解決しようとする課題】上記従来提案されている半導体外部記憶装置におけるデータ保証制御方式によれば、転送データの完結性を保証することができる。しかし、システム記憶に対するアクセス回数が増加するため、トランザクション処理におけるログデータのバッファなどをシステム記憶上に配置して、そのデータ保証を行った場合、システム記憶のアクセスによるCPUオーバーヘッドが増加するという問題がある。

【0013】ところで、完結性を保証しようとするデータが、例えば8バイトというような短いものであれば、中央処理装置が持つシステム記憶へのアクセス命令の中に、データ転送の完結性を保証する命令（例えば、CDSTS命令）があるので、その命令を用いることにより、障害時にもそのデータ単位については不定データにならないようにすることができる。しかし、ログデータなどは一般に長くなるのが普通であるので、1命令によって完結したデータの転送を行うことはできない。

【0014】同じように、ログデータの完結性を検査するための制御データについても、一般に8バイト以上の長さになるので、データ転送の完結性が保証される1命令によって書き込みを完結させることはできない。したがって、従来技術によるデータの完結性を保証する方式

(3)

特開平6-149485

4

では、データの退避等が必要になり、CPUのオーバーヘッドが増加する原因となっていた。

【0015】本発明は上記問題点の解決を図り、制御データおよびログデータなどの比較的長い転送データの完結性を効率よく保証できるようにし、CPUのオーバーヘッドを削減することを目的としている。

【0016】

【課題を解決するための手段】図1は、本発明の原理説明図である。図1において、10は半導体外部記憶装置等によって構成される不揮発性記憶装置、11は最新データ制御域、12Aは第1制御データ域、12Bは第2制御データ域、13はログバッファなどのデータ格納域、14は中央処理装置（CPU）およびメモリなどからなる計算機本体を表す。

【0017】計算機本体14におけるCPUは、例えば8バイトというような所定のデータ単位について、不揮発性記憶装置10への転送データの完結性を保証する命令の実行機構を有する。

【0018】本発明では、データ格納域13への転送データの完結性を保証するため、その検査に必要な制御データ（例えば転送アドレスや未転送データ長など）を制御データ域に設定する。さらに、8バイトより長い制御データ域への転送データの完結性を保証するため、本発明では次のようにする。

【0019】まず、不揮発性記憶装置10上に、制御データ域として、同じ構造を持つ第1制御データ域12Aと第2制御データ域12Bの2つを用意する。また、これらの制御データ域のうち、データ転送の完結した最新のデータを保持するほうをポイントする最新データ制御域11を不揮発性記憶装置10上に設ける。

【0020】制御データ域12Aまたは12Bを更新する場合に、最新データ制御域11からポイントされないほうの制御データ域に更新するデータ群を書き込み、そのデータ群の書き込みが完了した時点で、最新データ制御域11におけるポイントを、書き込みが完了した制御データ域のほうをポイントするように、データ転送の完結性が保証される1命令で書き込む。そして、最新データ制御域11からポイントされる制御データ域12Aまたは12Bを、有効なデータを保持する制御データ域として扱う。

【0021】データ格納域13へのデータ転送の完結性の保証、すなわちログデータなどの書き込み処理におけるデータ完結性の保証は、次のように行う。

(1) データ格納域13へのデータ書き込み要求を受け付ける。

【0022】(2) 最新データ制御域11を読み込み、第1制御データ域12Aと第2制御データ域12Bのどちらをポイントしているかにより、現在、信頼すべきデータが格納されている制御データ域を判断する。

【0023】(3) 最新データ制御域11がポイントして

(4)

特開平6-149485

5

いない空きの制御データ域側に、データ書き込み先の転送アドレスと、そのデータ長（未転送データ長）を書き込む。

【0024】(4) 要求されたデータを、データ格納域13に書き込む。

(5) 処理(3)で設定した未転送データ長をゼロクリアし、初期化する。

(6) 最新データ制御域11のポインタを、今まで空きであった制御データ域側をポイントするように切り替える。

【0025】データ読み込み処理におけるデータ完結性保証処理は、次のように行うことができる。

(1) データ格納域13からのデータ読み込み要求を受け付ける。

【0026】(2) 最初に、最新データ制御域11を計算機本体14のメモリに読み込む。

(3) 制御データ域12Aおよび制御データ域12Bを読み、未転送データ長が初期化されているかどうかを調べる。

【0027】(4) 未転送データ長が初期化されている場合、前のデータ転送は正常に行われているので、指定されたデータの読み込み処理を実行し、要求元へ返却する。

(5) 未転送データ長が初期化されていない場合、すなわち0でない場合、データ転送中に障害が起きたことが予想されるので、仕掛かり中のデータを再度データ格納域13に書き込み、以前のデータ転送を完結させる。その後、指定されたデータの読み込み処理を実行する。

【0028】

【作用】本発明では、不揮発性記憶装置10のデータ格納域13へ転送されたデータの完結性を検査できる情報を制御データとして持つ。この制御データ自身の完結性を保証するため、同じ構造を持つ制御データ域12A、12Bを2つ用意し、最新データ制御域11によって交互に切り替えて使用する。最新データ制御域11におけるポインタの切り替えは、データ完結性が保証される書き込み命令、例えば8バイト長の特殊書き込み命令（CDSTS命令など）を使用する。

【0029】これによって、制御データ域12A、12Bの制御データが、例えば複数の命令またはデータ完結性が保証されない命令によって更新される場合にも、それが、最新データ制御域11からの完結性が保証されたポインタによって有効化されることになるので、制御データ域12A、12Bのいずれかの制御データ全体のデータ完結性につながるようになる。

【0030】さらに、データ格納域13への転送データの完結性は、制御データ域12A、12Bの制御データによって検査可能であり、もし完結性が保証されない場合には、例えばトランザクション処理におけるバックアウトログファイル（BOF）などからのリカバリによ

6

て、転送処理途中のデータの復旧を行い、転送したデータの完結性を保証することができる。

【0031】

【実施例】図2は、本発明の実施例で用いるハードウェア構成例を示す。本発明は、一般の単一システム、または図2に示すような複合システムに適用することができる。

【0032】半導体で構成される不揮発性記憶装置10には、一般のメモリ部30と図1に示すデータ格納域13などの転送データロギング域31が設けられる。各計算機本体14A、14Bは、CPU33およびローカルに使用するメモリ部35を持ち、不揮発性記憶装置10に対しては、メモリ制御装置（MCU）32を介して、アクセスできるようになっている。また、転送データロギング域31への転送データを一時的に保持する転送データバッファ34を持つ。

【0033】図3は、本発明の適用システム説明図である。本発明の実施例を説明するに先立ち、その前提となる適用システムの概要を、図3に従って説明する。図3において、40はアプリケーション空間、41はアプリケーション空間40上で動作する応用プログラム、42はデータベースへの入出力のためのページバッファ、43はバックアウトログファイル（以下、BOFという）への入出力のためのBOFバッファ、44は応用プログラム41の処理に伴うログデータの格納に使用するタスクログバッファ、45はデータベース、46はトランザクション途中における異常発生時にデータベース45の内容を復元するためのデータが格納されるBOF、47はHLF（Historical Log File）への入出力のためのHLFバッファ群からなるHLFバッファプール、48はシステムで共用される共用空間、49はログデータの書き込みのためのバッファ、50はデータベース管理システム（DBMS）が動作するDBMS空間、51はデータベース45の格納媒体破壊時にそれを修復するためのログデータが格納されるHLF、52はHLF51への書き込みのためのバッファを表す。

【0034】以下の説明における処理(a)～(d)は、図3に示す(a)～(d)に対応している。

(a) トランザクションがスタートし、データベース45からのデータの読み込みを行う応用プログラム41は、読み込み（GET）マクロ命令を発行する。これにより、次の処理(a1)～(a6)が行われる。

【0035】(a1)読み込みのためのページバッファ42が満杯の場合、ページバッファ42に空きを作るため、次の(a4)までの処理を実行する。ページバッファ42に空きがある場合には、処理(a5)へ進む。

【0036】(a2)BOFバッファ43内の更新前データをBOF46に格納する。そのため、①BOF46の空きブロック検索処理、②更新前データの空きブロックへの転送処理、③BOF46のインデックス（IX）部の

(5)

特開平6-149485

7

更新処理を実行する。

【0037】(a3)タスクログバッファ44に更新後データを格納する。

(a4)データベース45の途中実更新を実行する。

(a5)ページバッファ42に、指定されたデータベース45のページを読み込む。

【0038】(a6)応用プログラム41内のレコード域へページバッファ45のレコードを転送する。

(b) 応用プログラム41は、データベース45を更新する場合、更新(MODIFY)マクロ命令を発行する。このとき、次の処理(b1)~(b4)が行われる。

【0039】(b1)レコード域のレコードを更新する。

(b2)ページバッファ42上の更新前データを、BOFバッファ43に転送する。

【0040】(b3)応用プログラム41内のレコード域から、更新後データをページバッファ42に転送する。

(b4)ページバッファ42上の更新後データのアドレスをHLFグループ毎に分割して、タスクログバッファ44に転送する。

【0041】(c) トランザクションを終了させる場合、応用プログラム41は、トランザクション終了(TRNEND)マクロ命令を発行する。これにより、次の処理(c1)~(c7)が実行される。

【0042】(c1)更新後データを共用空間48のバッファ49に転送する。

(c2)複合システムの場合、いわゆるグローバル・ロックを獲得し、他クラスタとの排他をとる(単独システムでは不要)。

【0043】(c3)共用空間48上のバッファ49にある更新後データを、BOF46に格納する。一般に、BOF46には更新前データだけが格納されるのが普通であるが、本実施例では、トランザクション処理中にデータベース45の格納媒体障害が発生したときの復旧処理を高速化するため、更新後データについてもBOF46に採取している。この格納処理のため、①BOF46の空きブロック検索処理、②更新後データの空きブロックへの転送処理、③BOF46のインデックス(IX)部の更新処理を実行する。

【0044】(c4)次に、共用空間48上のバッファ49にある更新後データを、不揮発性記憶装置10に設けられたHLFバッファプール47における該当するHLFバッファに転送し、排他を解除する。

【0045】(c5)その後、データベース45の一括実更新を実行する。

(c6)トランザクションの終了により、BOF46のインデックス部を初期化する。

【0046】(c7)トランザクションが獲得していた資源の占有を解除する(図示省略)。

(d)以上の処理(a)~(c)はトランザクション同期処理であるが、これに対し、トランザクション処理とは非同

8

期に、不揮発性記憶装置10上のHLFバッファが満杯になったことを契機にして、更新後データをバッファ52を経由しHLF51に格納する。

【0047】本発明は、例えば図3におけるHLFバッファプール47へのデータ転送における完結性保証のために用いることができる。図4はその本発明の実施例説明図、図5ないし図7はその実施例における詳細なデータ関連図である。

【0048】システム記憶である不揮発性記憶装置10は、計算機本体14からの8バイトのデータ完結性を保証する読み込み・書き込み命令によるアクセスが可能となっている。その命令の詳細についてはよく知られているので、ここでの詳しい説明は省略する。

【0049】最新データ制御域11、第1制御データ域12A、第2制御データ域12Bは、システム記憶に設けられ、図5に示すような情報を管理する。すなわち、最新データ制御域11は、最新の有効となっている制御データ格納域(CTBL)12Aまたは12Bへのアドレスを持つ。

【0050】HLFバッファプール47もシステム記憶内に設けられる。HLFバッファプール47の各エントリを、HLFバッファと呼ぶ。最新データ制御域11のエントリおよび第1制御データ域12A、第2制御データ域12Bは、各HLFバッファに対応して複数組設けられる。

【0051】各制御データ域12A、12Bにおけるアクセスキーは、対応するHLFバッファを識別するためのキー情報である。HLFバッファ内相対アドレスは、HLFバッファにおける空き領域の先頭アドレスを指しており、HLFバッファに対するレコード転送先のアドレスとなる。すなわち、次の更新後データは、このアドレスを先頭として転送される。未転送レコード長は、HLFバッファに転送しようとする更新後データの長さを指している。この未転送レコード長は、更新後データのHLFバッファへの転送が完了した時点で0に初期化される。未転送レコード長が0に初期化されている場合、そのデータ転送は完結していることを表す。

【0052】HLFバッファプール47の管理のために、図6に示すようなHLFバッファ管理テーブル70がシステム記憶内に設けられる。HLFバッファ管理テーブル70は、共通の管理情報を持つヘッダ部と、各HLFバッファ対応に設けられる個別部からなる。

【0053】ヘッダ部は、バッファ書出し処理中表示、使用可能なHLFバッファを示すエントリのアドレス、使用済み(使用中)HLFバッファを示すエントリのアドレスなどの情報を持つ。各HLFバッファ対応の個別部は、次のエントリをポイントするエントリ番号、HLFバッファプール47におけるHLFバッファへのアドレス、書出し責任元システム名、第1制御データ格納域の情報および第2制御データ格納域の情報などを持つ。

(6)

特開平6-149485

9

10

【0054】図4に示すBOF(BackOut log File)46は、トランザクション途中の更新前/更新後ログデータを格納している。これらのログデータは、トランザクションが完結すると初期化される。BOF46は、図7(A)に示すようなBOFインデックス部と、図7(B)に示すようなBOFデータ部とからなる。

【0055】BOFインデックス部は、ログデータが格納されたブロックおよびトランザクション情報を管理している。トランザクション情報としては、ログデータ発生システム名、ログデータ発生トランザクション識別子、ログデータが出力されたときのトランザクションの状態などの情報がある。

【0056】BOFデータ部は、ログデータを格納している部分であり、BOFデータ部のブロック長は、HLFバッファ長に等しい。ログデータの転送要求の際に指定されるパラメタバケット62は、図7(C)に示すような情報を持つ。すなわち、パラメタバケット62は、読み込みか書き込みかを示す要求コード、HLFバッファに対するアクセスキー、転送アドレス(転送先のHLFバッファ内相対アドレス)、転送データ長、ログデータが格納されているシステム記憶(SSU)転送元バッファ63のアドレスなどの情報を持つ。

【0057】図4に示すシステムにおいて、HLFバッファプール47へのログデータ書き込み処理は、以下のように行われる。

(1) 応用プログラム41から、トランザクション終了(TRNEND)マクロ命令が発行されると、データベース管理サブシステム(図示省略)を経由して、更新ログデータの保証処理が開始される。

【0058】(2) 図6に示すHLFバッファ管理テーブル70を読み込み、使用可能なHLFバッファ(これをカレントHLFバッファという)のアドレスを知る。

(3) システム起動後の最初のログデータの書き込み処理かどうかを判定し、最初でない場合には、処理(13)へ進む。

【0059】(4) 最初のログデータの書き込み処理である場合、最新データ制御域11を読み込み、最新の制御データ域のアドレスを知る(説明を簡単にするため、ここでは、第1制御データ域12Aが最新であるとする)。

【0060】(5) 最新ではない第2制御データ域12Bのエントリを読み込み、未転送レコード長が設定されているかどうかを判定する。未転送レコード長が初期化状態であれば、処理(13)へ進む。

【0061】(6) 未転送レコード長が設定状態であれば、第2制御データ域12BのHLFバッファ内相対アドレスを読み込み、HLFバッファにおける書き込み済みデータの最終アドレス(これをP1とする)を得る。

【0062】(7) 使用中のカレントHLFバッファの0～P1に存在するレコード制御部を読み込み、転送済み

トランザクション識別子を知る。

(8) BOFインデックス部を全検索する。転送済みでないトランザクション識別子が存在しない場合、処理(11)へ進む。

【0063】(9) 転送済みでないトランザクション識別子がBOFインデックス部中に存在するとき、そのトランザクション識別子の更新後データ格納ブロック番号を読み込む。

【0064】(10) 処理(9)で読み込んだブロック番号のブロックに格納された更新後データを、再度HLFバッファの相対アドレスP1に書き込む。(11) 第2制御データ域12Bの未転送レコード長を初期化する。

【0065】(12) 最新データ制御域11の制御データ域アドレスを、第2制御データ域12Bへ切り替える。

(13) 最新データ制御域11を読み込み、最新の制御データ域アドレスを知る(以下、説明を簡単にするため、最新の制御データ域は第2制御データ域12Bであるとする)。

【0066】(14) 第2制御データ域12Bのエントリを読み込み、当該HLFバッファのどこから使用可能であるかを知る。

(15) 以下の処理を行った後で、ログデータ保証のためのログ書き込みマクロ命令を発行する。

【0067】①主記憶のSSU転送元バッファ63に更新後データを書き込む。

②パラメタバケット62に必要な値を設定する。

(16) データ完結性保証機構60のログ書き込み処理では、パラメタバケット62の転送データ長を読み、書き込むレコード長を未転送レコード長として、最新データ制御域11が示していないほうの制御データ域(すなわち、第1制御データ域12A)に書き込む。

【0068】(17) パラメタバケット62の転送アドレスおよびSSU転送元バッファアドレスを読み込み、SSU転送元バッファ63の内容を、HLFバッファへ転送する。

【0069】(18) 第1制御データ域12Aの未転送レコード長を初期化する。

(19) 最新データ制御域11の制御データ域アドレスを、第1制御データ域12Aへ切り替える。

【0070】(20) ログ書き込みマクロ命令の発行元へ復帰する。

(21) データベース管理サブシステムを経由して、応用プログラム41へ復帰する。

【0071】複合システムにおけるダウン監視の延長で行われるHLFバッファ復旧処理は、以下のように行われる。

(1) ダウン監視機構(図示省略)から、他システムのダウン通知を受ける。このとき、ダウンシステムのトランザクションが占有していたレコードは、占有されたままの状態である。

(7)

特開平6-149485

11

【0072】(2) HLFバッファ管理テーブル70を読み込み、使用中のカレントHLFバッファアドレスを知る。

(3) 最新データ制御域11を読み込み、最新の制御データ域アドレスを知る(以下、説明を簡単にするため、最新の制御データ域は第1制御データ域12Aであるとする)。

【0073】(4) 第2制御データ域12Bのエントリを読み込み、未転送レコード長が設定されているかどうかを判定する。未転送レコード長が初期化状態であれば、
10 処理(12)へ進む。

【0074】(5) 未転送レコード長が設定されていれば、第2制御データ域12BのHLFバッファ内相対アドレスを読み込み、HLFバッファにおける書き込み済みデータの最終アドレス(これをP1とする)を得る。

【0075】(6) 使用中のカレントHLFバッファの0～P1に存在するレコード制御部を読み込み、転送済みトランザクション識別子を知る。

(7) BOFインデックス部を全検索する。転送済みでないトランザクション識別子が存在しない場合、処理(10)
20 へ進む。

【0076】(8) 転送済みでないトランザクション識別子がBOFインデックス部中に存在するとき、そのトランザクション識別子の更新後データ格納ブロック番号を読み込む。

【0077】(9) 処理(8)で読み込んだブロック番号のブロックに格納された更新後データを、再度HLFバッファの相対アドレスP1に書き込む。

(10) 第2制御データ域12Bの未転送レコード長を初期化する。
30

【0078】(11) 最新データ制御域11の制御データ域アドレスを、第2制御データ域12Bへ切り替える。

(12) ダウン監視機構へ復帰する。

【0079】なお、以上の実施例におけるデータ保証処理の一部を、ファームウェアによって実現することも可能である。

【0080】

【発明の効果】従来のデータ保証制御方式によるデータ転送では、不揮発性記憶装置に対するアクセス回数が増

12

加するため、トランザクション処理におけるCPUオーバーヘッドを削減できなかった。しかし、本発明によれば、ログバッファなどのトランザクション処理で頻繁に更新し、かつデータの完結性が要求されるデータを、半導体外部記憶装置で構成される不揮発性記憶装置に配置し、少ないCPUオーバーヘッドでアクセスすることができるようになる。その結果、以下の効果が得られる。

【0081】① データの転送中に、システムダウンなどの原因で転送処理が中断しても、転送データの完結性を保証することが可能になる。

② さらに、従来のデータ完結性保証処理に比べて、システム記憶に対するアクセス回数が少なくなるため、ログバッファなどのアクセス回数が多いデータに対して、本方法を適用することにより、高性能なシステムを構築することが可能になる。

【0082】その結果として、半導体外部記憶装置にデータベースおよびログファイルを配置し、これらをシステム間で共用するようなロードシェアシステムおよび高速ホットスタンバイシステムを構築することが可能になる。

【図面の簡単な説明】

【図1】本発明の原理説明図である。

【図2】本発明の実施例で用いるハードウェア構成例を示す図である。

【図3】本発明の適用システム説明図である。

【図4】本発明の実施例説明図である。

【図5】本発明の実施例に係るデータ関連図である。

【図6】本発明の実施例に係るデータ関連図である。

【図7】本発明の実施例に係るデータ関連図である。

【図8】従来技術の説明図である。

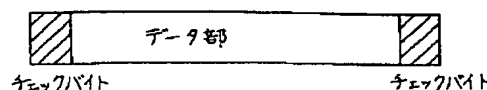
【図9】従来技術の問題点説明図である。

【符号の説明】

- 10 不揮発性記憶装置
- 11 最新データ制御域
- 12A 第1制御データ域
- 12B 第2制御データ域
- 13 データ格納域
- 14 計算機本体

【図8】

従来技術の説明図

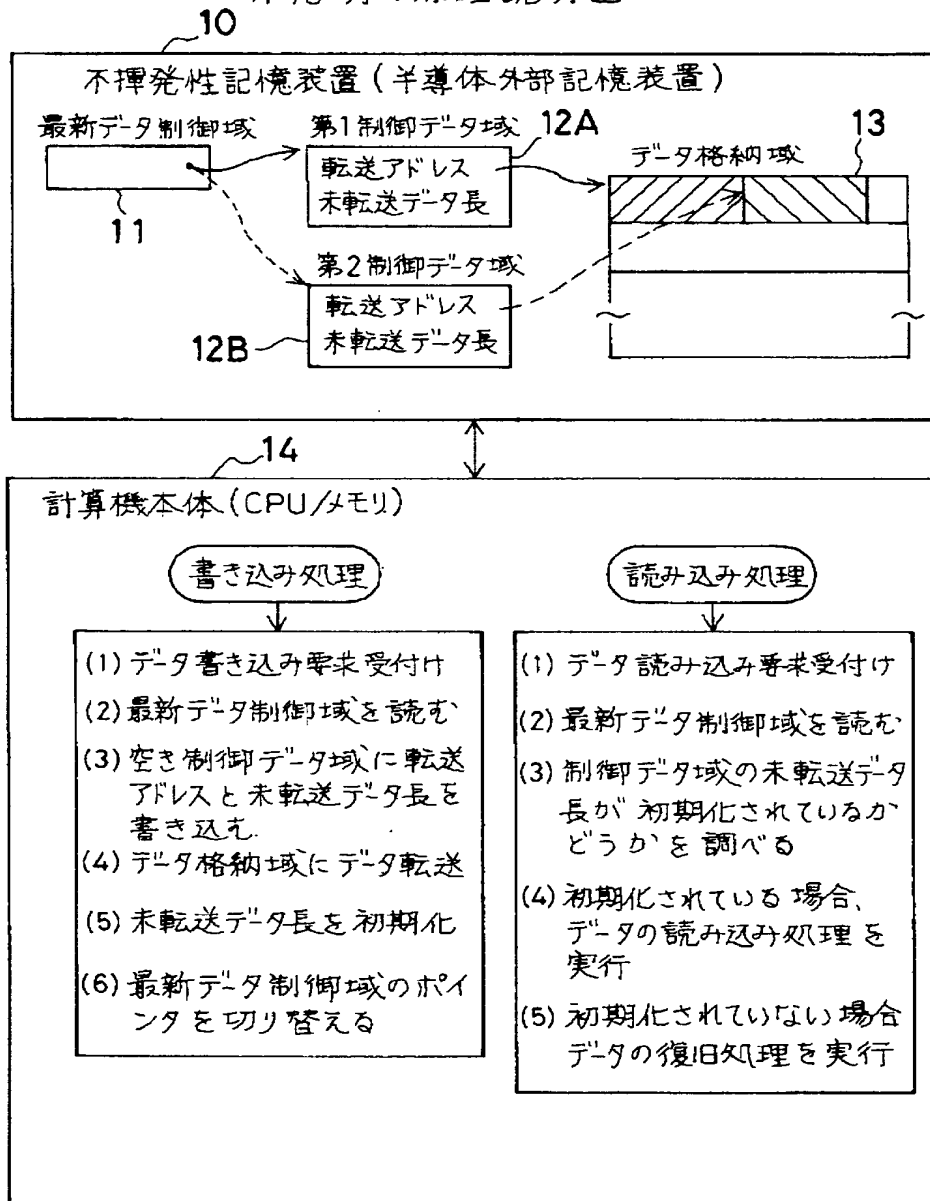


(8)

特開平6-149485

【図1】

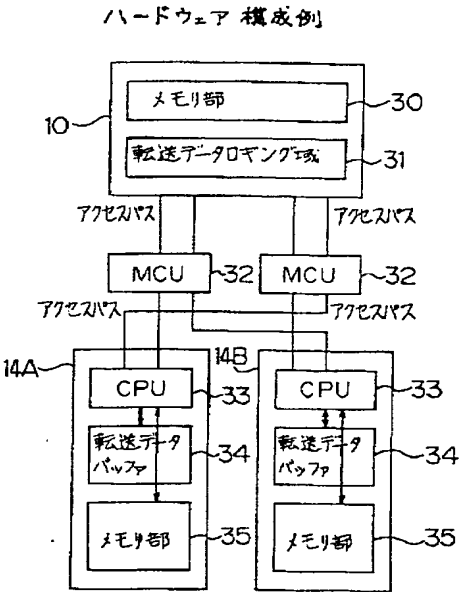
本発明の原理説明図



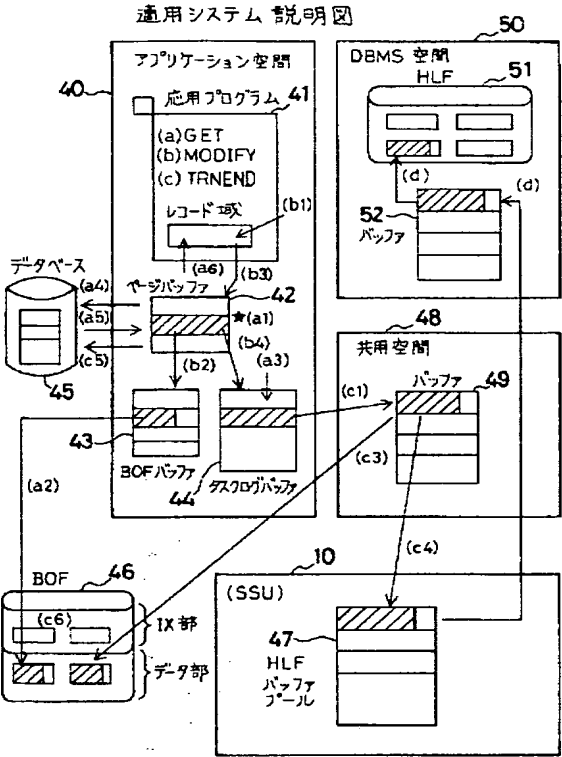
(9)

特開平6-149485

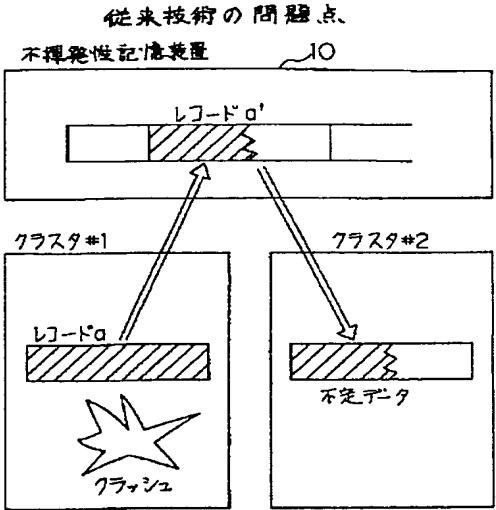
【図2】



【図3】



【図9】

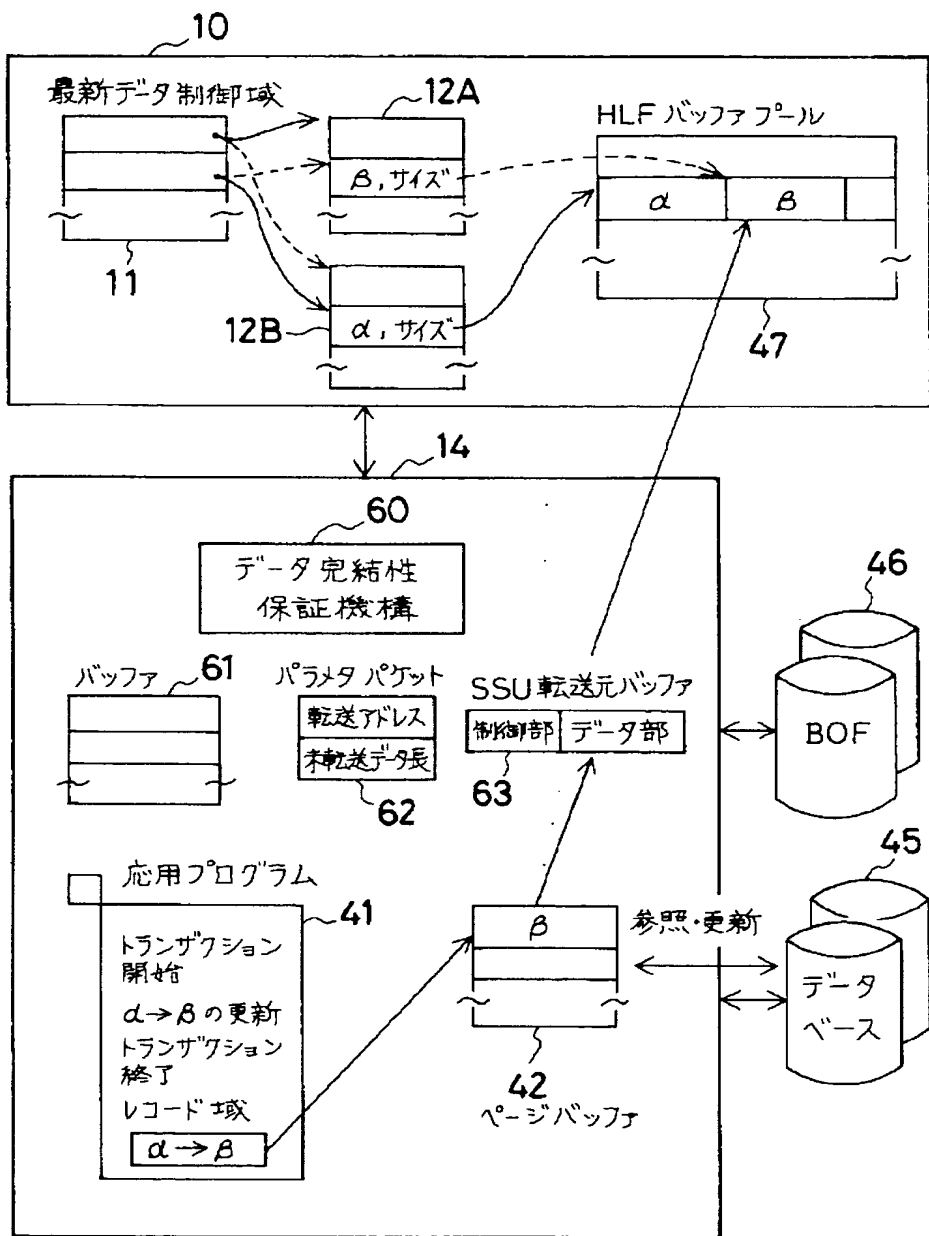


(10)

特開平6-149485

【図4】

本発明の実施例説明図

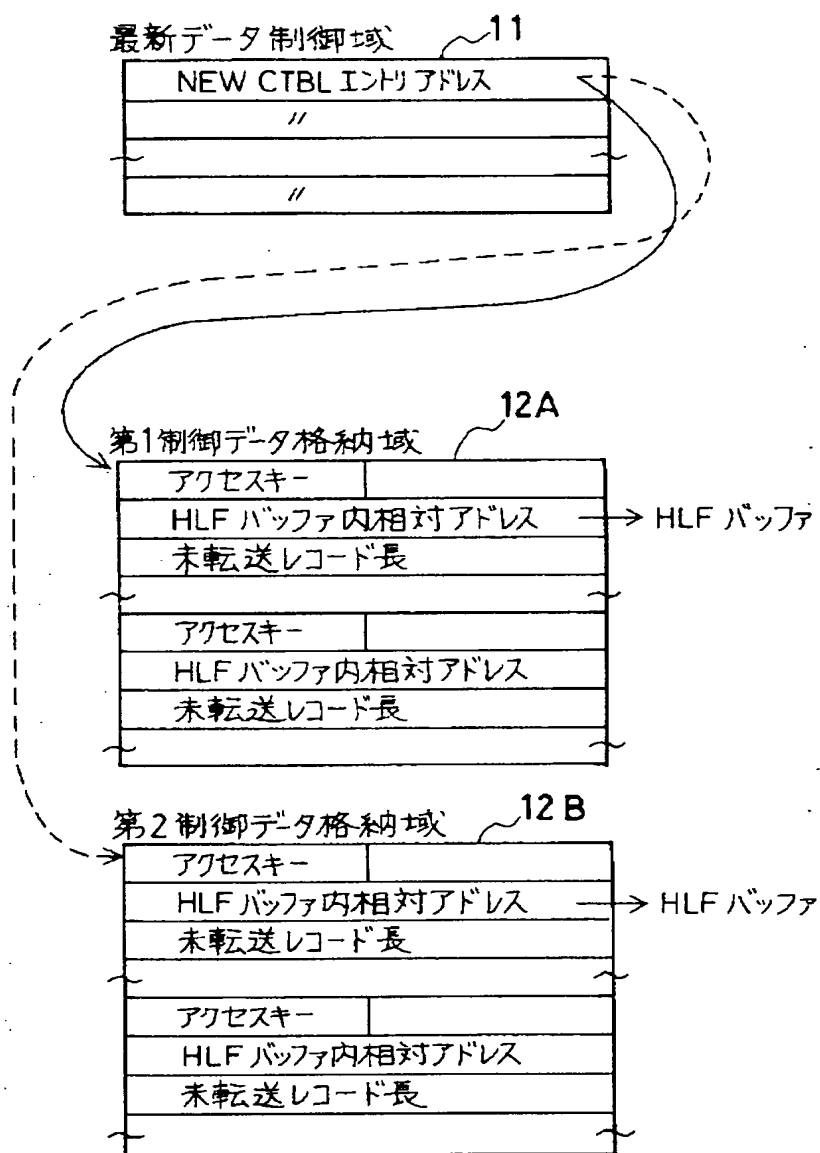


(11)

特開平6-149485

【図5】

データ関連図

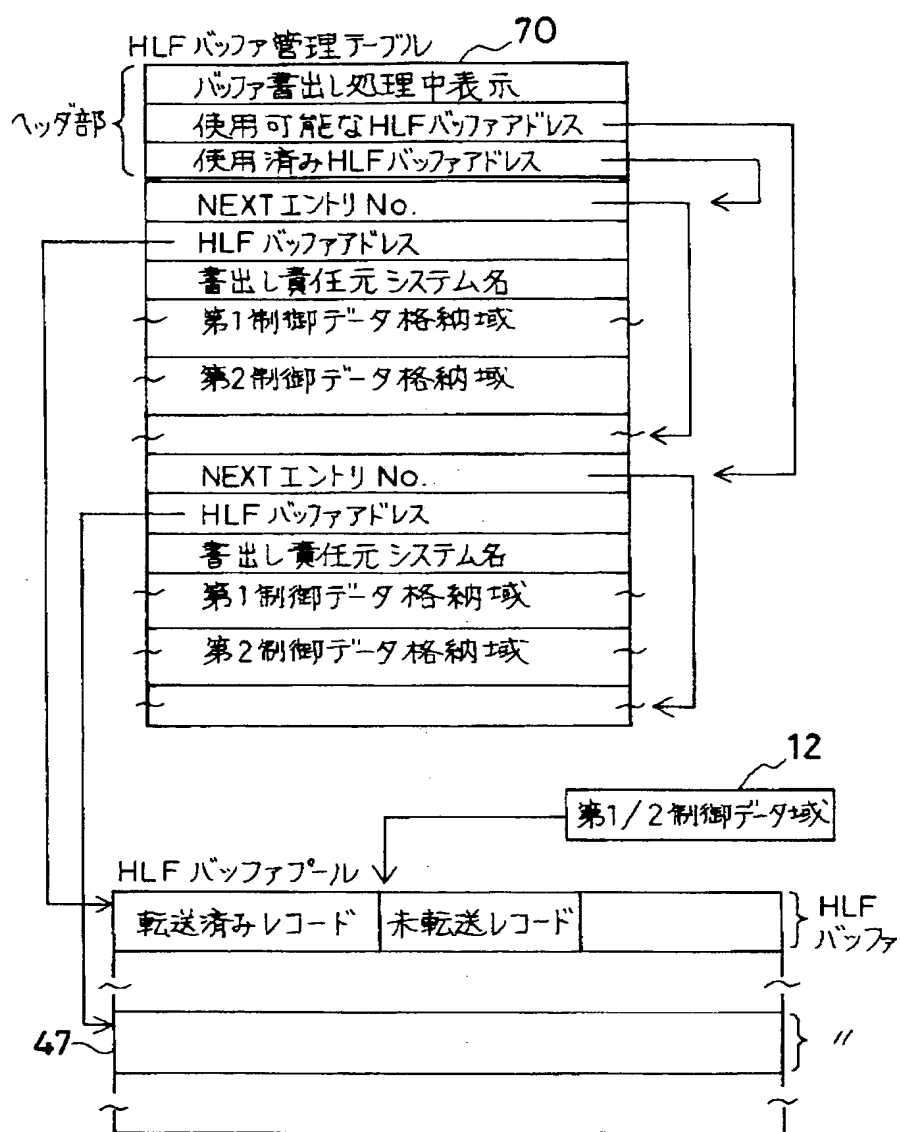


(12)

特開平6-149485

【図6】

データ関連図

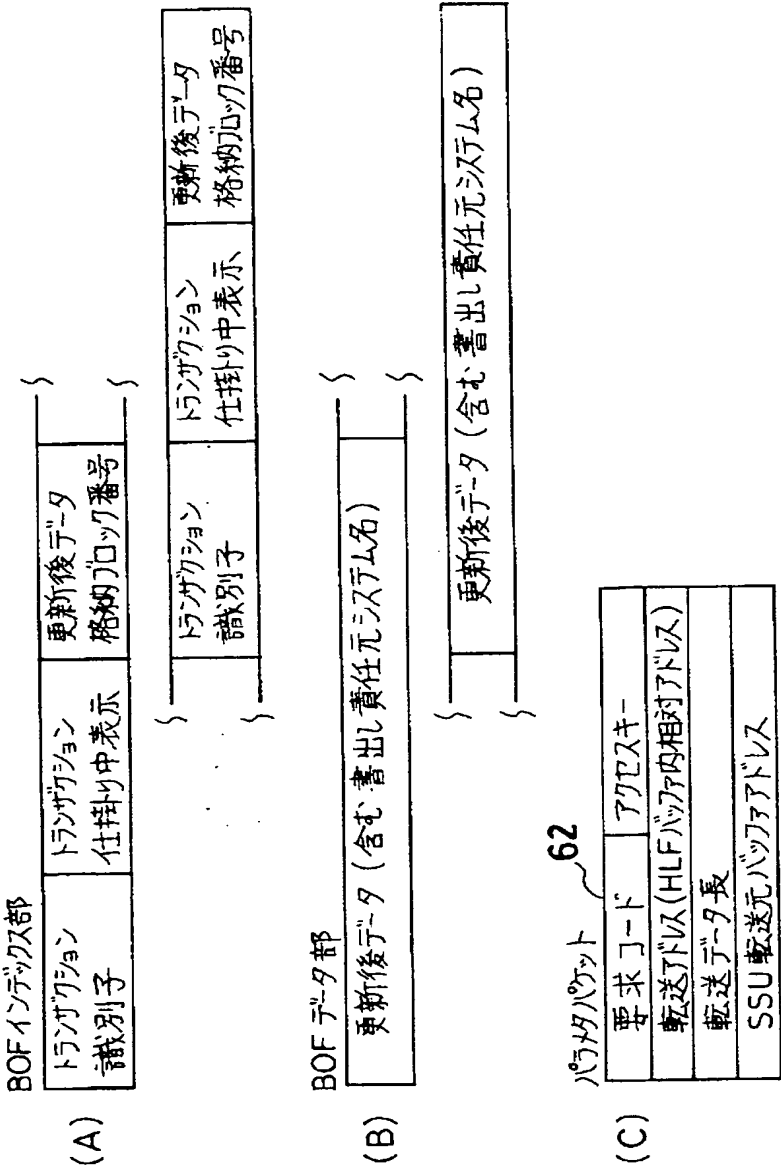


(13)

特開平6-149485

【図7】

データ関連図



This page blank (uspto)